Department of Pesticide Regulation

Paul E. Helliker
Director

MEMORANDUM

Gray Davis
Governor

Winston H. Hickox
Secretary, California Environmental Protection Agency

TO:	Exposure Assessment and Mitigation Program Staff Worker Health and Safety Branch	**HSM-02037**
VIA:	Joseph P. Frank, Senior Toxicologist Exposure Assessment and Mitigation Program Worker Health and Safety Branch	
FROM:	Sally Powell, Senior Environmental Research Scientist [original signed by S. Powell] Exposure Assessment and Mitigation Program Worker Health and Safety Branch	
DATE:	September 27, 2002	
SUBJECT**:**	APPROXIMATING CONFIDENCE LIMITS FOR UPPER BOUND AND MEAN EXPOSURE ESTIMATES FROM THE PESTICIDE HANDLERS EXPOSURE DATABASE (PHED V1.1)	

Background

The Worker Health and Safety Branch (WHS) conducts exposure assessments following guidelines that recommend which exposure statistics be used to estimate short- and intermediate-to-long-term exposures. These recommendations have recently been modified. The new recommendations presented a problem when the Pesticide Handlers Exposure Database (PHED, 1995) was used, because the recommended statistics could not be calculated.

This memorandum explains why the recommended statistics cannot be calculated for exposure estimates from the PHED and presents a method for approximating them. It gives a simplified method for calculating the approximate values and describes how to apply it in exposure assessments. (The method will be included in the revised version of the WHS exposure assessment guidance document (Thongsinthusak *et al*., 1993).)

In this section, the recommended exposure statistics and the reasons for adopting them are explained.

Recommended short-term exposure statistics

The 95th percentile of absorbed daily dosage (ADD) is generally used to represent short-term (up to 7 days in duration) exposure. The recommended statistic is the estimate of the 95th percentile of a lognormal population. When ADD is estimated using the PHED, the 90% upper confidence limit on the 95th percentile should be used. (These recommendations apply only to point estimates of exposure.)

Reasons for the recommendations

With these short durations, concern is for the highest exposure an individual may realistically experience while performing a label-permitted activity. WHS generally uses the 95th percentile of daily exposure as this "upper bound". The reason a high percentile is estimated rather than the maximum itself is that, in theory, the maximum value of a lognormal population can be indefinitely large. In practice, exposure values must be bounded because a finite amount of active ingredient (AI) is applied. The 95th, rather than a higher percentile, is estimated because the higher a percentile, the less reliably it can be estimated, as well as because the lognormal distribution may not be a perfect description of the distribution of exposure values, especially at the upper extremes. A parametric estimate† of the population percentile is used, rather than the sample percentile, because upper-end sample percentiles, in samples of the sizes usually available to exposure assessors, are both statistically unstable and known to underestimate the population values. The parametric estimate, on the other hand, is more stable, being based on all the observations rather than a single value; moreover, it is adjusted, in effect, for sample size, correcting some of the underestimation bias due to small samples

When data from the PHED are used to estimate exposure, an upper confidence limit on the percentile should be used in place of the percentile itself, in order to increase confidence in the estimate by accounting for some of the uncertainty added by using surrogate data whose relevance to the target exposure scenario cannot be fully assessed. The 90% confidence level is used by statistical convention.

Recommended intermediate- to long-term exposure statistics

The arithmetic mean of absorbed daily dosage (ADD) is recommended to represent exposures of longer than 7-day duration. When ADD is estimated using the PHED, the 90% upper confidence limit on the arithmetic mean should be used. (These recommendations apply only to point estimates of exposure.)

Reasons for the recommendations

The reason for using average daily exposure rather than an upper bound is that over these durations, a worker is expected to encounter a range of daily exposures. That is, with increased exposure duration, repeated daily exposure at the upper-bound level is unlikely. WHS uses the arithmetic mean rather than the geometric mean or the median because, although it can be argued that the latter statistics better indicate the location of the center of a skewed distribution, it is not the center that is of interest in exposure assessment, but the *expected magnitude* of exposure. While extremely high daily exposures are low-probability events, they do occur, and the arithmetic mean appropriately gives them weight in proportion to their probability. In contrast, the geometric mean gives decreasing weight as the value of the exposure increases, and the

† A parametric estimate is one that uses assumptions about the distribution of the parent population in addition to the observed data values. For example, the sample 95th percentile is the value that is ≥ 95% of the observations, while the parametric estimate assuming lognormality is $\exp\left\{\hat{\mu} + t\left(0.95, n-1\right)\hat{\sigma}\right\}$.

median gives no weight to extreme exposures. When the PHED is used to estimate exposure, an upper confidence limit on the arithmetic mean is used, to increase the confidence in the estimate by accounting for some of the uncertainty added by using surrogate data whose relevance to the target exposure scenario cannot be fully assessed. The 90% confidence level is used by statistical convention. (In most instances, the mean daily exposure of individuals over time is not known. However, the mean daily exposure of a group of persons observed in a short-term study is believed to be the best available estimate of the mean for an individual over a longer period.)

Why exact confidence limits cannot be calculated for PHED dermal exposure estimates

Calculating confidence limits requires estimates of the mean and standard deviation. (Consider the confidence interval for the mean of a normal distribution: $\hat{\mu} \pm \hat{\sigma}/\sqrt{n} \cdot t_{(1-\alpha/2,n-1)}$). When *dermal* exposure is estimated using the PHED, its standard deviation is not available. The reason is that most of the studies included in PHED V1.1 used patch dosimetry to measure dermal exposure. The studies differed widely in the numbers of patches used and the body regions on which patches were placed. PHED V1.1 calculates the mean and coefficient of variation for each body region using all available patches for the region, so body-region exposures are usually based on different studies and different numbers of observations. PHED calculates mean total exposure as the sum of the body-region means. The standard deviation of a sum can be calculated from the standard deviations of the summed elements and their intercorrelations. However, the correlations among body regions in PHED are not known and cannot be estimated readily since each region is based on different observations. Thus, the standard deviation of total dermal exposure cannot be calculated.

Assumptions about the distribution of total dermal exposure

The standard deviation of total dermal exposure cannot be calculated from the PHED output, but by making certain assumptions, it is possible to find approximate confidence limits for the mean and 95th percentile. It is assumed that:

1) total dermal exposure is lognormally distributed, and
2) total dermal exposure has a coefficient of variation (CV) of 100 percent.

Lognormality is assumed because many exposure datasets are characterized well by lognormal distributions. A CV of 100% is in the range seen for individual body regions in PHED subsets.

Calculation of approximate confidence limits

For the convenience of WHS exposure assessors, a simplified method was developed for obtaining the confidence limits. The basis of the method is given here. (Definitions of the statistical symbols used are given at the end of this memo.)

Confidence limit for the 95th percentile

Confidence limits on percentiles, usually called tolerance limits, are described in Hahn and Meeker (1991; Section 4.6). The 90-percent upper confidence limit for the 95th percentile of a lognormal distribution is

$$\exp\left\{\hat{\mu} + g'(0.90;0.95;n)\cdot\hat{\sigma}\right\}.$$

The symbols $\hat{\mu}$ and $\hat{\sigma}$ represent the arithmetic mean and standard deviation of the natural logarithms of the data. Values of the multiplier $g'_{(1-\alpha,\, p,\, n)}$, which depend on the confidence level (1- α), the percentile (p) and the sample size (n), are tabled in Hahn and Meeker (1991; Table A12.d).

The arithmetic mean of a lognormal distribution is $\exp\left\{\mu + \tfrac{1}{2}\sigma^2\right\}$. An estimate of the mean, based on a set of data, is $\exp\left\{\hat{\mu} + \tfrac{1}{2}\hat{\sigma}^2\right\}$.

The ratio of the confidence limit to the estimated arithmetic mean,

$$\frac{\exp\left\{\hat{\mu} + g'(0.90;0.95;n)\cdot\hat{\sigma}\right\}}{\exp\left\{\hat{\mu} + \tfrac{1}{2}\hat{\sigma}^2\right\}} = \exp\left\{g'(0.90;0.95;n)\hat{\sigma} - \tfrac{1}{2}\hat{\sigma}^2\right\},$$

does not depend on the value of the mean[‡].

Further, for any lognormal distribution with CV = 100%, the standard deviation, σ, of the corresponding normal distribution equals 0.83255. This follows from $CV_{\text{lognormal}} = \sqrt{\exp(\sigma^2) - 1}$ (Crow and Shimizu, 1988, p. 10).

Therefore, the 90-percent upper confidence limit (UCL) for the 95th percentile of a lognormal distribution with CV = 100% can be estimated from a sample of size n as

$$\text{UCL} = \text{sample arithmetic mean} \times \exp\left\{g'_{(0.90;0.95;n)}\cdot 0.83255 - \tfrac{1}{2}(0.83255)^2\right\}.$$

The multiplier $\exp\left\{g'_{(0.90;0.95;n)}\cdot 0.83255 - \tfrac{1}{2}(0.83255)^2\right\}$ depends only on the sample size n. Values of the multiplier for various n are given in Table 1.

‡ See technical endnote.

Table 1. 90% upper confidence limit for the 95th percentile as a multiple of the mean (CV of 100% assumed).

n	Confidence limit as a multiple of the mean	n	Confidence limit as a multiple of the mean	n	Confidence limit as a multiple of the mean
3	58.86	15	4.916	27	4.093
4	19.06	16	4.794	28	4.059
5	11.99	17	4.688	29	4.025
6	9.278	18	4.599	30	3.995
7	7.868	19	4.516	35	3.868
8	7.003	20	4.445	40	3.769
9	6.422	21	4.379	50	3.631
10	5.998	22	4.321	60	3.535
11	5.682	23	4.267	120	3.274
12	5.428	24	4.218	240	3.112
13	5.224	25	4.172	480	3.005
14	5.057	26	4.131		

In order to avoid giving the impression of greater numeric accuracy than this method can really provide, these multipliers have been rounded to the nearest whole number in Table 2. This table will appear in the revised exposure assessment guidance document. The inclusion in Table 2 of *n* as small as 5 is not meant to endorse the use of such small samples. Minimum sample sizes for PHED will be discussed in the revised guidance document.

Table 2. Short-term exposure estimate (90% upper confidence limit for the 95*th* percentile) as a multiple of PHED arithmetic mean exposure [a].

n	Multiplier	n	Multiplier
5	12	9 - 11	6
6	9	12 - 19	5
7	8	20 - 119	4
8	7	≥ 120	3

[a] The exposure estimate is calculated by multiplying the ADD based on arithmetic mean total dermal or inhalation exposure by the multiplier corresponding to the median number of observations.

Confidence limit for the arithmetic mean

Similarly, the ratio of a given confidence limit for the mean to the mean is constant for fixed σ and n. The upper 1- α percent confidence limit for the mean of a lognormal distribution is given by

$$\exp\left\{\hat{\mu} + \tfrac{1}{2}\hat{\sigma}^2 + \frac{\hat{\sigma}}{\sqrt{n-1}} \cdot C\left(\hat{\sigma};n-1;1-\alpha\right)\right\},$$

where the values of C are tabled in Land (1975) and can also be obtained from a computer program written by Land *et al.* (1987). The ratio of the 90-percent confidence limit to the mean for a lognormal distribution with CV = 100% is

$$\exp\left\{\frac{0.83255}{\sqrt{n-1}} \cdot C\left(0.83255;n-1;0.90\right)\right\}$$

The Land *et al.* computer program (1987) was used to obtain the ratio of confidence limit to mean by sample size (Table 3).

Table 3. 90% upper confidence limit on the arithmetic mean as a multiple of the mean (CV of 100% assumed).

n	Confidence limit as a multiple of the mean	n	Confidence limit as a multiple of the mean	n	Confidence limit as a multiple of the mean
3	18.309	20	1.396	90	1.150
4	4.499	25	1.337	100	1.141
5	2.924	30	1.297	120	1.127
6	2.372	35	1.268	140	1.117
7	2.091	40	1.246	180	1.102
8	1.920	45	1.228	200	1.096
9	1.804	50	1.214	240	1.087
10	1.719	55	1.201	300	1.077
11	1.655	60	1.191	480	1.060
12	1.603	65	1.182	600	1.053
13	1.561	70	1.174		
14	1.526	75	1.167		
15	1.497	80	1.161		

The rounded multipliers (Table 4) will appear in the revised exposure assessment guidance document. The inclusion in Table 4 of *n* as small as 5 is not meant to endorse the use of such small samples. Minimum sample sizes for PHED will be discussed in the revised guidance document.

Table 4. Intermediate- to long-term exposure estimate (90% upper confidence limit for the mean) as a multiple of PHED arithmetic mean exposure [a].

n	Multiplier
5	3
6-14	2
≥ 15	1

[a] The exposure estimate is calculated by multiplying the ADD based on arithmetic mean total dermal or inhalation exposure by the multiplier corresponding to the median number of observations.

Applying the method

To use this method in an exposure assessment:

1 - Obtain from the PHED output the arithmetic mean of total dermal exposure in µg /lb ai, the arithmetic mean of inhalation exposure in µg /lb ai, and the median number of observations for all body regions *plus* inhalation. (That is, one median *n* will apply to both exposures.)

2 - Calculate ADD for dermal exposure and for inhalation exposure.

3 – Find the multipliers in Tables 2 and 4 (or the equivalent tables in the revised exposure assessment guidance document) corresponding to the median *n.*

4 - Multiply both dermal ADD and inhalation ADD by the multiplier from Table 2 to get the short-term exposure estimates.

5 - Multiply both dermal ADD and inhalation ADD by the multiplier from Table 4 to get the intermediate- or long-term exposure estimates.

6 - The estimate of short-term total (dermal and inhalation) exposure is the sum of the separate short-term estimates for dermal and inhalation.[§]

7 - The estimate of long-term total (dermal and inhalation) exposure is the sum of the separate long-term estimates for dermal and inhalation.[§]

[§] It is not true in general that the confidence (or tolerance) limit of a sum is equal to the sum of the confidence (or tolerance) limits. In this special case, however, because dermal and inhalation exposures are treated as independent and both are assumed to have CV=100%, the sum of limits and the limit of sums are approximately equal. Summing the limits appears to overestimate the true tolerance limit by about 0-13% (mean 2%) and the true confidence limit by 0-4% (mean 1%) in PHED datasets.

Statistical symbols

σ standard deviation of the normal distribution of the logs of a lognormally distributed variable

$\hat{\sigma}$ sample estimate of σ

μ arithmetic mean of the normal distribution of the logs of a lognormally distributed variable

$\hat{\mu}$ sample estimate of μ

$\exp\{x\} = e^x$, the exponential function, or antilog of the natural logarithm

$g'_{(1-\alpha, p, n)}$ statistical multiplier whose value depends on a confidence level $1-\alpha$, a percentile p and sample size n

References

Crow, E. L. and K. Shimizu (1988). Lognormal Distributions: Theory and Applications. New York, Marcel Dekker, Inc.

Hahn, G.J., and Meeker, W.Q. 1991. *Statistical Intervals: A Guide for Practitioners*. New York, John Wiley & Sons, Inc.

Land, C.E. 1975. Tables of confidence limits for linear functions of the normal mean and variance, in Selected Tables in Mathematical Statistics, Vol.3. American Mathematical Society, Providence, R.I., pp. 385-419.

Land, C.E., L.M. Greenberg, C. Hall and C.C. Drzyzgula. 1987. BTNCTD: Exact confidence limits for arbitrary linear functions of the normal mean and variance. Unpublished computer program. Radiation Epidemiology Branch, National Cancer Center, NIH.

PHED. 1995. Pesticide Handlers Exposure Database, Version 1.1. Prepared for the PHED Task Force: Health Canada, U.S. EPA and the American Crop Protection Association, by Versar, Inc., Springfield, VA.

Thongsinthusak, T., Ross, J. H., and Meinders, D. 1993 (currently under revision). Guidance for the preparation of human pesticide exposure assessment documents. HS-1612. Sacramento, CA: Worker Health and Safety Branch, Department of Pesticide Regulation, California Environmental Protection Agency.

Technical note

The argument presented on pages 4 and 6 for the constancy of the ratio of a confidence limit to the estimated mean, e.g.,

$$\frac{\exp\left\{\hat{\mu} + g'(0.90;0.95;n) \cdot \hat{\sigma}\right\}}{\exp\left\{\hat{\mu} + \tfrac{1}{2}\hat{\sigma}^2\right\}} = \exp\left\{g'(0.90;0.95;n)\,\hat{\sigma} - \tfrac{1}{2}\hat{\sigma}^2\right\},$$

does not strictly apply to the sample mean. The sample arithmetic mean is not identical to $\exp\left\{\hat{\mu} + \tfrac{1}{2}\hat{\sigma}^2\right\}$. Both are valid estimators of the population mean, but the exponentiated quantity has a small upward bias, while the sample mean is unbiased. The result, when the multiplier is applied to the sample mean, is that the upper confidence limit is underestimated. Based on simulations, the underestimation appears to be no greater than about 3 percent of the correct value of the confidence limit. The underestimation is greatest in extremely small samples (n = 3) and decreases as n increases.